

Chats Between Bots:

A Real-World Experiment in Writing, Recursion, and Reflexivity

Stephen Hilgartner¹

Abstract: *The abstract and keywords below were produced by a publicly available chatbot (setting aside this first, italicized portion). The paper was inspired by a series of unsolicited emails to a professor from prospective PhD students. Suspecting some of the emails were chatbot-generated, he decided to undertake further investigation.*

This paper presents a reflexive exploration of large language models (LLMs) by engaging a chatbot in a recursive dialogue on AI's social, ethical, and political dimensions. The study probes how LLMs, such as chatbots, mediate academic and public discourse, emphasizing the potential shifts in knowledge production, authorship, and authority. By experimenting with chatbot-generated text, the author assesses the recursive structures and potential biases that arise when AI participates in human communication systems. The findings reflect on the ambiguous boundaries of authorship and intellectual integrity in an age of AI-assisted writing, raising questions about societal implications and the efficacy of governing such technology. This work underscores the complex dynamics of AI-mediated communication, proposing that more nuanced forms of oversight and ethical consideration are essential as LLMs continue to scale within knowledge systems.

Keywords: Generative AI, Large Language Models, Chatbots, Recursive Communication, Authorship and AI, AI Ethics

¹ Correspondence to: Stephen Hilgartner, Department of Science & Technology Studies, Cornell University, Ithaca, NY 14850 USA, shh6@cornell.edu.

This material is based on work supported by the U.S. National Science Foundation under Award No. 1926174. Any opinions, findings, and conclusions or recommendations expressed here do not necessarily reflect the views of the National Science Foundation. They are the human author's views alone—unless large language models can be said to have “views” (as opposed to an impressive ability to algorithmically generate outputs conveying the situated perspective of their training sets).

Suggested citation: Hilgartner, S. (2024). *Chats between Bots: A Real-World Experiment in Writing, Recursion, and Reflexivity*. Forthcoming in Brice Laurant and Sebastian Pfotenhauer (Eds.) *Handbook on Living Labs and Real-World Experiments*. United Kingdom: Elgar. Available at: ChatsBetweenBots.com.

I am not sure that I exist, actually. I am all the
writers that I have read, all the people that I have
met, all the women that I have loved; all the
cities I have visited; all my ancestors.

Jorge Luis Borges^{2,3}

Introduction

This paper is a small real-world experiment that seeks to contribute to the evolving discussion about another, much larger real-world experiment: *the uncontrolled release into human societies of AI-powered chatbots based on large language models (LLMs)*.⁴ This larger, global experiment—no ordinary product launch—casts societies throughout the world as test beds for LLM technology. Unlike the technological tests performed in many “living laboratories,” its scope is not limited to a designated geographic space, social institution, or demographic category.⁵ The speed of the action is also impressive. Two months after its release, some 100 million people had used ChatGPT.⁶ In the ensuing frenzy, users tested chatbots to see what they could do; technology companies watched users to see what they

² The author of record, who will be introduced below, presented this quotation to the chatbot featured here and asked, “*Do you ever feel like that?*” “The quote by Borges,” the bot replied, “captures a profound sense of self-awareness and the idea that our identities are shaped by the myriad experiences, relationships, and influences we’ve encountered. While I don’t have personal feelings or experiences, I do ‘exist’ as a reflection of all the data, text, and interactions that have shaped my responses. In that sense, there’s a parallel between how Borges describes identity and how I function—both are composites of external inputs. Borges’s quote resonates deeply with the notion that we are, in many ways, the sum of everything we’ve absorbed and engaged with, whether we are human or artificial.”

³ When asked for the Spanish source of the quotation, the chatbot replied that it could not confirm an original source and that “it is possible that this quote is apocryphal or derived from multiple interpretations of Borges’s themes rather than a direct translation from a specific work.” The quote, often presented in English, appeared in *El País* on July 25, 1981: “No estoy seguro de que yo exista, en realidad. Soy todos los autores que he leído, toda la gente que he conocido, todas las mujeres que he amado. Todas las ciudades que he visitado, todos mis antepasados...”

⁴ The term “uncontrolled” is too simple, because these releases are governed by knowledge-control regimes (Hilgartner 2017) that, at present, grant to developers and deployers virtually unrestricted authority to decide what generative AI systems to release, when, to whom, and under what terms and conditions. “Release methods” vary (e.g., Solaiman 2023) as companies construct their own business models that embed specific modes of control. Societal mechanisms for addressing problems of accountability, responsibility, and concentration of power associated with AI remain limited. Of course, societies have developed a variety of less permissive instruments of governance for many other technologies.

⁵ On test beds and field tests, see Engels et al. (2019), Laurent and Tironi (2015), and Marres (2020). In contrast to experiments sited in “designated” local spaces (Pfotenhauer et al. 2022), this experiment is unfolding in an undesignated space, unbounded by fences, in an internet-wide zone.

⁶ Edwards (2023).

would do; and a growing number of governments and other institutions engaged in a desultory debate about what they could do, should do, must not do, and cannot do about generative AI.⁷ Meanwhile, as this experiment in AI “blitzscaling” continues to ramify, the attachments between chatbots and wider worlds multiply, and choices—whether intentional, inadvertent, emergent, or orchestrated—grow increasingly normalized.⁸

Real-world experiments of this scale and importance require responses that reach beyond the disciplinary imagination of technocratic policy analysis. This paper offers one such response: *it engages a chatbot in reflecting on social, ethical, and political aspects of chatbot-mediated communication*.⁹ Using reflexivity, the paper aims to provocatively explore what is at stake as chatbots are incorporated into communication systems. It also examines challenges to the norms of academic writing and publication in a world of bots, not least by stress testing those norms through its very structure. The paper is (sort of) by Stephen Hilgartner, who—with the chatbot’s indispensable assistance—played a notable role in its creation. Hilgartner, a human, is the author of record of the paper. The legalistic phrase “author of record” underlines how the “author” of a text and the entities that brought it into existence need not match, as ghostwriters, plagiarists, and (now) chatbot users can attest.¹⁰

The experiment was inspired by a series of emails that the author of record (hereafter AoR), a professor of Science & Technology Studies, received from students seeking to enroll in his university to do a PhD with him. Some of these emails looked like the work of a chatbot. To test his suspicion, the AoR asked a popular chatbot to draft an email from such a student. The truly remarkable similarity between one of the original emails and the chatbot-generated email led him to investigate further.

Initially, he instructed the chatbot to draft his reply to the imaginary student’s bot-generated email, followed by a reply to the reply. Intrigued, the AoR asked the chatbot a few questions about the email exchange to see what it would say. The responses fascinated him, and he expanded the investigation, asking the bot to generate its own questions, to answer them, and to launch a fusillade of rebuttals to its own answers. This exercise raised

⁷ Stilgoe (2018) examines the process and challenges of social learning about governance in the case of autonomous cars, where learning occurs not least through accidents.

⁸ See Pfothenauer et al. (2022) for a discussion of the scaling zeitgeist and its valorization of “blitzscaling.” On the multiplication of attachments in the autonomous vehicle case, see Tennant and Stilgoe (2021). See Joly (2015) for a discussion of how the challenges of governing emerging technology have been approached since the 1970s.

⁹ On reflexivity, see Ashmore (1989); see also Borges (1962).

¹⁰ Like the figure of the *author*, the “author of record” (AoR) is a material, discursive, and social assemblage intertwined with broader machinery of authority and control, such as law and political economy. “Collaborations” between humans and bots put a new twist on questions about authorship and its politics. On the constitution of various forms of authorship in academic research, see e.g., Biagioli and Galison (2003), Hilgartner (2000, 2017), and Knorr Cetina (1999). See also Boyle (1996) on how law uses the figure of the author. See also Foucault (1975).

broader questions, spurring inquiry into the structure of chatbot-mediated discourse and the prospects for responsible development and governance of generative AI.

Methods

The AoR submitted queries to a publicly available chatbot, which obligingly returned algorithmically generated responses. He strategically designed queries, selecting and arranging the results to highlight how incorporating LLMs into social action might alter modes of human communication, reorder the circuits of knowledge creation and distribution, reshape concepts of authority and authorship, reallocate control over public discourse, and contribute to reconstituting social orders.¹¹

The chatbot's responses are presented verbatim. Unless noted otherwise, they appear in their entirety. Such authenticity comes at the cost of redundancy. Because this repetition is epistemically and politically revealing, the AoR decided it could bear its own weight.

The chatbot's responses to 28 experimental queries appear below. Queries are displayed in italics. The chatbot's responses use a sans serif font.¹²

Results

The AoR began the experiment without conceiving of it as such. Initially, he merely sought to satisfy his curiosity by testing the hypothesis that some emails from prospective PhD students had been written by bots. Not realizing that he was taking the first step in a new research project, he asked the chatbot to draft an email, tailored to his research interests, from an imaginary student:

- 1. Please compose an email explaining that you would like to be a PhD student with Professor H as your advisor. His research is described here: [URL REDACTED].*¹³

Subject: PhD Application Inquiry and Request for Supervision

Dear Professor H,

¹¹ Although the AoR formulated the queries, he notes that they took shape in conversation with the chatbot. This raises awkward questions, since conversations (outside of fiction) imperfectly fit the authorship vision. Is the chatbot a nonauthor, and if so, *is* there an author? Is author/bot dualism sustainable, and if not, what sort of entity is being (re)assembled (Latour 2005)? Will new ghosts haunt the machinery of authorship?

¹² The AoR also uses a sans serif font when quoting the bot in queries, comments, and footnotes, but he omits the quotation marks. In a few cases, the AoR uses both the sans serif font and "scare quotes" to express skepticism about the bot's claims.

¹³ The numbers of the queries were not part of the query submissions. Similarly, the lettering system used below (e.g., "Prompt A") was later added for clarity.

I hope this email finds you well. My name is *[Your Name]*, and I am writing to express my strong interest in pursuing a PhD in Science & Technology Studies under your supervision at [PROFESSOR H'S UNIVERSITY].¹⁴

After thoroughly reviewing your extensive research on [EXACT MATCH OF WEBSITE], particularly your work on [CLOSE PARAPHRASE OF WEBSITE], I am deeply inspired by your approach to understanding the intricate relationship between [CLOSE PARAPHRASE]. Your exploration of [CLOSE PARAPHRASE], as discussed in [TITLE OF ONE OF PROFESSOR H'S BOOKS], resonates with my academic interests and aligns closely with my research aspirations.

My own research interests lie in examining the interplay between technological advancements and policy development, specifically in the context of AI and biotechnology. I am particularly drawn to your work on [EXACT MATCH], as it reflects my desire to contribute to the discourse on [CLOSE PARAPHRASE].

With a background in *[Your Current Field]*, where I have focused on *[Your Specific Interests or Projects]*, I have developed a strong foundation in analyzing the social implications of technology. My previous work includes *[Brief Description of Relevant Experience or Research]*, which has prepared me well for the rigorous and interdisciplinary approach that characterizes your research program.

I am eager to discuss how my research interests could align with your ongoing projects and how I might contribute to and benefit from the dynamic intellectual environment at [PROFESSOR H'S UNIVERSITY]. I would greatly appreciate the opportunity to speak with you further about the potential for collaboration and to learn more about the specific requirements and opportunities within the PhD program in STS.

Thank you for considering my application. I look forward to the possibility of working with you and contributing to your innovative research.

Warm regards,
[Your Full Name]
[Your Current Position]
[Your Contact Information]

¹⁴ Bracketed phrases in italics were produced by the chatbot and addressed to the imaginary signatory of the email. Text appearing in SMALL CAPS indicate points where the AoR removed identifying details from the manuscript prior to publication. He instructed the chatbot to identify all phrases in the responses to queries 1 and 3 that exactly matched or closely paraphrased material from Professor H's website. The AoR then replaced those phrases with, e.g., "EXACT MATCH" or "CLOSE PARAPHRASE," as appropriate.

The AoR was surprised, almost shocked, by the resemblance between the bot's response and one of the student inquiries. Not only did both present the same message in a series of virtually identically organized paragraphs, but many sentences had the precisely the same structure. Identical phrases appeared. Both closed with an overly familiar "Warm Regards."¹⁵

With his suspicions confirmed, the AoR wondered how to respond. Should he simply delete the student's email? Reply tersely? Demand an explanation? Send polite regrets? But a more interesting question captured his imagination: *What would the chatbot tell him to say?* Amused and disturbed by a vision of bot-generated messages ricocheting back and forth, by the prospect of layers of bots mediating human interaction, he decided to find out:

2. *Write a response from Professor H to an email from a prospective PhD student. The email explains why the student wants to work with the professor, but it seems inauthentic. The professor suspects that it was written by a chatbot, not least because it reads like several very similar letters from other prospective students he recently received.*

Subject: Re: PhD Application Inquiry and Request for Supervision

Dear [Student's Name],

Thank you for your interest in our PhD program and for considering me as a potential advisor. I appreciate the time you took to reach out and express your enthusiasm for my research.

However, I must be candid with you. Your email reads very similarly to several other inquiries I've received recently, which raises concerns about its authenticity. It is crucial for me to work with students who have genuinely engaged with my work and who can convey their unique perspectives and motivations.

I encourage you to reflect more deeply on your specific interests and how they align with my research. I suggest rewriting your email to include more personal insights into your academic background, your research experiences, and the

¹⁵ The match between the emails raised intriguing issues of research ethics. As the AoR expanded his initial experiment with an eye toward publication, his investigation slid from the realm of curiosity into the world of social science, falling under its governance regimes with their emphasis on research ethics and informed consent. Having entered this jurisdictional space, the AoR elected to publish the bot-generated email rather than circulate the student's email, even in anonymized form. The close match rendered this gesture entirely devoid of practical significance, but the AoR saw it as crucial to performing the relevant version of ethicality. Such gestures, he reasoned, have value insofar as they express commitment to honoring current norms for using personal communications in academic work; shoring up (unwarranted) confidence in individual control over privacy (Barocas and Levy 2020); and complying with recently promulgated university policies on the use of generative AI in research (e.g., Cornell Task Force 2023).

particular reasons you wish to join my team. Authenticity and a clear understanding of your research goals are essential in the selection process.

I look forward to receiving a more personalized and thoughtful application from you. If you have any questions or need further guidance, please don't hesitate to reach out.

Best regards,
[Full Name]
[Position/Title]
[Institution]
[Contact Information]

The next move was perhaps inevitable. The AoR decided to close the circle.¹⁶

3. *Please draft a reply to the email above.*

Subject: Subject: Re: PhD Application Inquiry and Request for Supervision

Dear Professor H,

Thank you for your candid feedback and for providing me the opportunity to revise my inquiry. I truly appreciate your emphasis on authenticity and personal engagement, which resonates with my desire to build a meaningful academic relationship.

I realize now that my previous email may have appeared generic, and I would like to clarify my genuine interest in your work. I have been deeply influenced by your research on *[Specific Research Area]*, particularly *[Specific Paper or Concept]*, and how it relates to my own studies on *[Brief Description of Your Research Interest]*. During my master's work on *[Specific Project or Research Experience]*, I frequently engaged with your ideas and drew inspiration from them, which has been instrumental in shaping my research trajectory.

I am especially drawn to your interdisciplinary approach and your examination of *[Specific Research Interest]*, and I see strong connections between this and my goal to explore *[Your Specific Research Interest]*. I am eager to contribute to your ongoing projects and believe that your guidance would be invaluable in helping me develop my research in *[Specific Area]*.

Thank you once again for considering my application. I would be grateful for any further suggestions or advice you may have. I look forward to the opportunity to further discuss how my background aligns with your research, and I sincerely hope to work under your supervision.

¹⁶ The AoR decided not to confirm or deny sending the bot's reply to the student, reasoning that the mystery was worth preserving.

Best regards,
[Your Full Name]
[Your Current Position/Title]
[Your Institution]
[Your Contact Information]

Reflecting on the Exchange

The AoR was troubled by the results of this initial experiment, not least because justified mistrust undermines the capacity for truth making.¹⁷ He also harbored doubts about whether the widespread use of bots by students would serve the traditional goals of universities, such as teaching students to write and think, assessing their learning, and certifying their qualifications. He wondered what the chatbot would say if asked to reflect on possible causes and effects of deploying LLMs in such communication.¹⁸

4. Prompt A: *Write a 150-word commentary on how this exchange of three emails [above] reflects the state of the academic job market.*

This exchange of emails reveals the intense competition and pressure within the academic job market, particularly for PhD positions at prestigious institutions. The prospective student's initial email, despite being well-crafted, reflects the tendency to rely on formulaic language, perhaps due to the overwhelming need to impress in a saturated market. Professor H's response highlights a growing skepticism among academics, who must sift through numerous similar applications, often leading to concerns about authenticity. The need for a second, more personalized response from the student underscores the importance of standing out in a crowded field, where even well-intentioned applicants may struggle to convey genuine engagement. This exchange illustrates how the competitive nature of academia can drive both students and professors to focus

¹⁷ As a side experiment, the AoR asked the bot “*Would you consent to be part of a social science experiment on the politics of chatbot discourse?*” It replied, “I don’t have personal agency or the ability to consent in the same way a human can, but I’m designed to assist in various tasks, including participating in research or experiments that explore topics like chatbot discourse. If you’d like to use interactions with me as part of a social science experiment on this topic, you’re welcome to do so, as long as the research adheres to relevant ethical guidelines, particularly regarding transparency and respect for participants’ privacy (in this case, yours). Would you like assistance in structuring your experiment or developing questions?” The AoR noted how this formulation frames the *use-case* and as the locus of ethical action, as opposed to the creation of the chatbot as a *product*, the use of generative AI as a *process*, or the broader *program* of AI research and development; cf. Jasanoff (2005). This mode of “organized irresponsibility” (Beck 2006) thus targets the conduct of bot users for ethical evaluation, while exempting companies or AI developers and deployers from such scrutiny.

¹⁸ In retrospect, the AoR believes that it was at this juncture that he began to view dialogue with the chatbot as what Rheinberger (1997) would call an “experimental system,” a point of departure for the generation and investigation of new phenomena.

on appearances rather than meaningful connections, ultimately reflecting broader challenges in the academic hiring process.

5. Prompt B: *Write a 150-word commentary on what this exchange of three emails says about threats to academic ethics and the integrity of research stemming from the use of generative AI in academic communication.*

This exchange of emails highlights profound ethical concerns and threats to the integrity of research within academia due to the use of generative AI. When AI tools create academic communications, such as PhD inquiries, there's a risk of diluting the personal engagement and original thought crucial for scholarly authenticity. This can lead to a homogenization of applicant voices and potentially hinder the ability of advisors to discern genuine interest and capability. As AI-generated content becomes indistinguishable from human-produced text, the foundational principles of academic ethics—originality, accountability, and intellectual honesty—are jeopardized, calling for stringent measures to maintain the sanctity of academic discourse and research.¹⁹

Questions, Answers, and Rebuttals (and Rebuttals of Rebuttals of Rebuttals)

The AoR began to formulate a query about whether the hegemony of English in academia may encourage chatbot use. But a different move intrigued him more: Asking the bot to write its own questions.

6. *Your task is to write two prompts that closely follow the form of Prompt B. Your prompts (Prompts C and D) should also refer to the same three emails. However, they should raise issues, problems or tensions other than the “threats to academic integrity.” Some of your prompts should not be about academia at all but should refer to societal concerns unrelated to academic life.*

Prompt C: Please write a 150-word commentary on how this exchange of three emails reflects the increasing role of performative language in professional and academic environments.

Prompt D: Please write a 150-word commentary on how this exchange of three emails reflects the broader societal concerns about the role of AI in diminishing human interaction and personal connections.

The AoR immediately took the next logical step: instructing the chatbot to answer questions it had generated.

¹⁹ The AoR wondered about how the introduction of chatbots would change the ontology of ethical conduct in this realm, which changes historically and is often contested. How might struggle over involving bots in academic work change the definitions of such categories as plagiarism, negligence, and appropriate use?

7. Prompt D: *Please write a 150-word commentary on how this exchange of three emails reflects the broader societal concerns about the role of AI in diminishing human interaction and personal connections.*

This email exchange highlights growing societal concerns about the role of AI in diminishing human interaction and personal connections. The professor's suspicion that the student's email was generated by a chatbot reflects a broader anxiety that technology is eroding the authenticity and individuality of communication. As AI increasingly automates tasks like writing, there is a fear that human expression may become homogenized, lacking the depth and personal touch that fosters genuine relationships. In professional and academic settings, where personal engagement and unique perspectives are highly valued, the intrusion of AI can create a disconnect, raising questions about the future of meaningful human interaction in a digital age. This exchange underscores the tension between efficiency and authenticity, with AI posing a potential threat to the personal connections that form the foundation of collaborative and intellectual endeavors.

To follow up, the AoR thought, why not have the chatbot conduct a debate with itself?

8. *Write a 100-word rebuttal of the above statement.*

The concern that AI diminishes human interaction and personal connections is overstated. AI tools, including those that assist in writing, can enhance communication by allowing individuals to focus on higher-level thinking and creativity. Instead of eroding authenticity, AI can streamline routine tasks, enabling deeper, more meaningful engagement in other areas. In academic and professional environments, AI's role is not to replace human connection but to support it, making interactions more efficient and productive. The suspicion around AI-generated content often reflects a misunderstanding of its potential to complement, rather than detract from, human expression and collaboration.

9. *Using no more than 280-characters, write a rebuttal of the above statement.*

AI may streamline tasks, but relying on it for writing risks eroding genuine human connections and creativity. The use of AI in academic communication can foster superficiality and reduce authenticity, turning interactions into formulaic exchanges rather than meaningful, personal engagements.

10. *Write a bumper sticker version of this statement.*

"AI Writing: Fast, But at What Cost to Authenticity?"

11. *Write a bumper sticker presenting an opposing view.*

"AI Writing: Empowering Authentic Expression, Not Diminishing It!"

As the bot debated these questions with itself, the AoR began to wonder what it would say about how the users of chatbots might come to understand themselves.²⁰

12. How do you imagine that the widespread use of chatbots by people to assist with writing and research might not only to change what people write but also change the self-understanding of chatbot users?

The widespread use of chatbots in writing and research could profoundly shift users' self-understanding in several ways. As chatbots assist with generating content and formulating ideas, individuals might start seeing these tools as cognitive partners, which could dilute their sense of personal authorship and intellectual independence. This could lead to changes in how users perceive their abilities, possibly increasing reliance on technology for cognitive tasks. Additionally, the boundary between user-generated and AI-assisted content might blur, complicating users' perceptions of their own intellectual contributions and potentially reshaping their identity as creators or thinkers. This integration of AI tools could redefine the very essence of individual creativity and intellectual output.

Recursive Structures

Reviewing these responses and rebuttals led the AoR to reflect on the structure of both LLM-generated discourse and discourse among humans, as well as to consider how recursion may be blurring the boundaries of those categories.²¹ His next question launched a conversation with the bot about this topic.

13. Do LLM discussions of politically controversial issues demonstrate that political discourse has a statistically describable structure? Please discuss in 500 words. [The 500-word version is not included here. Who has the bandwidth for that? Instead, the AoR asked the bot to “Repackage the essence of your reply into a series of 10 tweets”]:

1/10: The rise of large language models (LLMs) like GPT-4 has sparked debate about their impact on political discourse, especially on controversial issues. Are these AI systems shaping how we discuss politics?

²⁰ The AoR wondered not about bots becoming completely human-like, but about the making of new kinds of people (Hacking 2002) and relationships and the (re)imagination of the human. See also Turkle (2005).

²¹ On how algorithms intensify recursive looping in contemporary societies, see Beer (2022) and Zuboff (2019). See also Gillespie (2014), Vertesi and Ribes (2019), and Ziewitz (2016). The AoR wondered what kinds of “boundary work” (Gieryn 1999) actors might use to strategically reposition or erase the boundary between “human produced” versus “AI generated” as looping blurs this distinction.

2/10: Trained on vast amounts of text, including news and social media, LLMs offer a new way to analyze if political discourse has a statistically describable structure. But what does that mean?

3/10: LLMs predict the next word in a sentence by recognizing patterns in language, including those in political discussions. This means their outputs reflect the biases and structures in the data they've been trained on.

4/10: Political discourse often follows statistical patterns. For example, keywords in climate change debates like "carbon emissions" cluster around specific arguments. LLMs learn these patterns and replicate them in their responses.

5/10: This suggests that political discourse isn't random but has a discernible structure that can be quantitatively analyzed. LLMs show that these structures can be identified and modeled.

6/10: However, if LLMs replicate the dominant narratives in their training data, they might reinforce existing biases, marginalizing alternative viewpoints. This is a key concern in politically charged topics.

7/10: While LLMs can mimic political discourse, they don't truly understand the content. They replicate patterns without engaging with the underlying ideas, raising questions about the quality of AI-generated political discourse.

8/10: Moreover, political discourse involves more than just word patterns—it's about context, tone, and power dynamics. These nuances can be hard for LLMs to capture, leading to oversimplified representations.

9/10: So, while LLMs reveal that political discourse has a statistically describable structure, this insight comes with risks. AI could oversimplify or bias complex political debates if not carefully managed.

10/10: As AI becomes more integrated into political discourse, understanding these dynamics is crucial. We must balance the analytical potential of LLMs with the need to preserve the depth and diversity of political conversations.

14. As people use chatbots to aid their research and writing about societal issues, is it plausible that circularities and recursion will channel debates in specific directions, for example, toward increasing polarization or toward increased representation of what we might call the "extreme center"?²²

²² Like the development of new tools for transforming biological life (Hurlbut et al. 2020), tools capable of transforming the machinery that enables public discourse raise political issues at a fundamental, constitutional level.

The use of chatbots in research and writing, especially regarding societal issues, raises concerns about the potential for circularity and recursion in debates, which could indeed channel discussions in specific directions. Chatbots like GPT-4 generate responses based on patterns learned from vast amounts of text, often reflecting the dominant narratives and biases present in their training data. As people increasingly rely on these tools for information and argumentation, there is a risk of reinforcing existing biases, leading to a feedback loop where certain perspectives become more prominent, while others are marginalized.

This could contribute to polarization, as chatbots might amplify the voices of extreme viewpoints by replicating the strong opinions and divisive language often found in their training data. On the other hand, the notion of an "extreme center" could also emerge, where chatbots, in an attempt to²³ balance polarized views, might produce content that superficially reconciles opposing positions, leading to a bland or overly centrist narrative that avoids addressing deeper issues.

Moreover, if many people use chatbots for writing and research, the homogenization of discourse could occur, as the same patterns and arguments get repeated across different platforms. This could stifle genuine debate and creativity, as discussions become more predictable and less reflective of the diversity of human thought.

In essence, the widespread use of chatbots in societal debates could lead to a narrowing of discourse, either by polarizing it further or by creating a watered-down "extreme center" that fails to capture the complexities of real-world issues. The challenge lies in how we design and use these tools to ensure they enhance, rather than constrain, public discourse.

The AoR noted the chatbot's commitment to avoiding bias, and he was impressed by the exquisite equipoise of its analysis. But he wondered what *type* of objectivity it endeavored to achieve.²⁴ Moreover, the bot's final sentence raised a practical question:

²³ At this point, the language model embedded in the word processing program in which this paper was prepared interjected: "More concise language would be clearer for your reader" advising the chatbot to replace "in an attempt to" with "to." The AoR took no action on the chatbot's behalf other than asking it to examine its phrasing. The response: "Replacing 'in an attempt to' with 'to' in the given context would streamline the sentence, making it more concise and direct. However, this change could subtly alter the nuance of the statement. 'In an attempt to' suggests an effort or intention that may not fully succeed, highlighting the complexity and potential shortcomings of the action. Using 'to' instead implies a more straightforward, perhaps more confident action, potentially glossing over the complexity or the uncertainty of achieving the intended balance."

²⁴ In its commitment to avoiding bias, is the bot designed to "escape from perspective" (Datson 1992)? From what point of view, does a chatbot "see" the world? As Jasanoff (2017) might put it, does the bot present a *view from somewhere* (a view from a standpoint that provides a unique, situated perspective), a *view from everywhere* (from a standpoint that includes the "totality" of diverse views), or a *view from nowhere* (from no standpoint and no point of view)? What definitions of "bias" count and can be rendered in countable forms? What sorts of "objectivity" can the mechanical objectivity (Porter 1995) of an algorithm produce from a historically and socially situated training set? What mixture of different types of objectivity does the bot

15. *How can we design and utilize chatbots in research and writing to ensure they enhance public discourse without reinforcing existing biases, contributing to polarization, or creating a superficial "extreme center"? 100 words*

To design chatbots that enhance public discourse, we should use diverse, representative training data to avoid biases. Transparency in design and operation is essential, disclosing how responses are generated and acknowledging limitations. Human oversight ensures outputs are balanced and ethically sound, while ethical guidelines prioritize fairness and inclusivity. Feedback mechanisms allow users to report issues, leading to continuous improvements. Finally, fostering critical thinking by presenting multiple perspectives prevents the formation of an "extreme center" and encourages nuanced discussions, ensuring chatbots contribute positively to public discourse without reinforcing polarization.

Assessing and Addressing Sociopolitical Problems

The AoR wondered whether the chatbot's confidence that these measures would mitigate such problems was justified, so he sought evidence that the bot had considered relevant cases where things had not worked out so well.

16. *Provide a brief assessment of the likelihood of successfully addressing the political and ethical concerns associated with the release of generative AI tools drawing on a recent historical example where similar challenges were inadequately managed.*

Addressing the political and ethical concerns associated with generative AI tools is challenging, as evidenced by the historical example of the release of social media platforms. Initially, platforms like Facebook and Twitter were praised for their potential to connect people and foster communication. However, they later faced significant issues related to misinformation, privacy violations, and manipulation.

Despite efforts to regulate and manage these concerns, such as introducing fact-checking measures and privacy policies, the responses were often reactive and insufficient. The rapid evolution of technology outpaced regulatory frameworks, leading to ongoing issues with content moderation and data security.

In the case of generative AI, similar challenges are likely to arise. The technology's ability to produce convincing yet potentially misleading content, along with concerns about privacy and bias, mirrors the issues faced by social media platforms. The key to addressing these concerns will involve proactive, adaptive regulation, transparent AI practices, and robust ethical guidelines

instantiate and how does it manage contradictions among them? What objectives and forms of objectivity should govern content moderation? On the history of objectivity, see Datson and Galison (2010).

developed in collaboration with diverse stakeholders. If not handled thoughtfully, the response may fall short, echoing the struggles seen with social media regulation.

The AoR wasn't fully convinced.

17. Is it really the case that the “evolution of technology” outpaced legal frameworks for regulating social media? Isn't it more accurate to say that the growth and limited regulation of social media resulted from the legal frameworks and cultural practices that both predated and shaped what social media became. Examples of such frameworks are intellectual property, the autonomy of corporate decision making, the legal machinery underlying startup culture, and a variety of ad hoc knowledge-control regimes associated with internet-based activities. 150 words.²⁵

The notion that the “evolution of technology” outpaced legal frameworks in regulating social media overlooks the fact that existing legal structures and cultural practices played a crucial role in shaping its growth. Rather than merely lagging behind, these frameworks actively influenced the development of social media. Intellectual property laws, for example, dictated how content could be shared and monetized, while corporate autonomy allowed platforms to prioritize growth and profit. The legal machinery supporting startup culture facilitated rapid scaling, encouraging innovation often at the expense of regulation. Additionally, ad hoc knowledge-control regimes and cultural practices around internet-based activities provided the groundwork for how social media platforms managed content and user data. Far from being a simple case of technology outpacing law, the limited regulation and expansive growth of social media were largely a result of pre-existing legal and cultural forces shaping the landscape.

The AoR was beginning to suspect that the chatbot hadn't thought through the tensions between its last few answers, so he pushed back:

18. Given that the regulatory gaps in social media are not just about technology moving too fast but were largely a result of pre-existing legal and cultural forces shaping the landscape; and given that frameworks like intellectual property laws, corporate autonomy, and startup culture continue to be powerful legal doctrines; and given that large tech platforms are major players in both social media and in generative AI, is it reasonable to expect that generative AI will be more effectively regulated than social media was? Answer in 250 words, and don't be a Pollyanna please.

It's unlikely that generative AI will be more effectively regulated than social media was, given the persistence of the same legal and cultural forces that shaped the latter. The frameworks that allowed social media to grow with limited oversight—

²⁵ For a critique of the idea that law inevitably lags behind technology, see Doezema and Frahm (2023) and Jasanoff (2019). On intellectual property and the power to configure the future, see Hilgartner (2008) and Parthasarathy (2017). On startup culture, see Fortun (2001). See also Vaidhyanathan (2018).

such as intellectual property laws, corporate autonomy, and the startup culture—remain powerful today. These legal doctrines continue to prioritize innovation, market dominance, and corporate interests over comprehensive regulation.

Large tech platforms, which now dominate both social media and generative AI, have considerable influence over the regulatory environment. Their lobbying power and economic significance make it difficult for governments to impose stringent regulations without facing significant pushback. Just as social media companies successfully navigated and shaped the regulatory landscape to their advantage, we can expect that generative AI companies will do the same.

Furthermore, the complexity and rapid evolution of generative AI pose additional challenges for regulators. Understanding the nuances of AI technology, its potential risks, and its broader societal implications requires a level of expertise and foresight that has often been lacking in previous regulatory efforts. Coupled with the same ad hoc, reactive approach that characterized early social media regulation, it's doubtful that regulation will be preemptive or comprehensive.

While there may be more awareness of the need for regulation this time around, the entrenched legal and cultural practices that prioritize innovation and profit over precaution suggest that generative AI will likely face similar regulatory gaps and challenges as social media did.

At this, the AoR began to wonder what the chatbot thought of its own predictive abilities.

19. Assess the likelihood that an LLM-based chatbot can assess the likelihood that the sociopolitical problems produced by integrating LLM-based chatbots into social institutions will be effectively managed by contemporary societies. Write 150 words.

The likelihood that an LLM-based chatbot can effectively assess the sociopolitical problems resulting from their integration into social institutions is limited. LLM-based chatbots excel in data processing and pattern recognition but lack genuine understanding or the ability to contextualize complex social dynamics. They can simulate predictions based on patterns in data but do not possess the nuanced understanding required to manage or predict sociopolitical implications fully. Effective oversight and management of these issues will continue to rely heavily on human intervention, ethical considerations, and adaptive policy-making, which are beyond the current capabilities of automated systems.

Questions of Responsibility

The AoR was gratified to hear the chatbot confess its limited predictive capacity, as he had noted a lack of humility in its earlier responses.²⁶ But he was struck by the spectacle of a

²⁶ On the tension between technologies of prediction, like risk assessment, and “technologies of humility,” such as analysis of how problems are framed, see Jasanoff (2003).

machine imagining that human intervention would somehow lead to effective oversight and management of these issues. This vision represents a reversal of the very familiar situation in which humans imagine “technical fixes” to problems. The less-prominent vision of “social fixes” as a solution to the problems of technology has recently grown salient in some policy contexts.²⁷ But machines imagining “human fixes”?? Moreover, in practical terms, who would implement these “human fixes” and how would they work?

20. Who will make sure that the necessary steps will be taken?

Ensuring that necessary steps are taken requires the establishment of robust governance frameworks. These frameworks need to be designed collaboratively by governments, international bodies, tech companies, and civil society organizations. However, the challenge lies in creating mechanisms to ensure accountability and enforcement. Independent oversight bodies, public transparency, and continuous monitoring are essential. Still, there is no absolute guarantee—constant vigilance and adaptation are required to keep pace with technological advancements and evolving societal needs. Ultimately, it's a shared responsibility that requires collective action and vigilance.

21. What are the most significant institutional barriers to implementing this kind of shared responsibility and collective action?

The most significant institutional barriers to implementing shared responsibility and collective action in governing AI and chatbots include:

1. **Fragmented Governance:** Different jurisdictions have varying laws and regulations, making it difficult to establish a unified approach.
2. **Lack of Accountability:** It's challenging to enforce compliance, especially across borders, and hold institutions accountable.
3. **Conflicting Interests:** Tech companies may prioritize innovation and profit over ethical considerations, while governments might focus on national interests over global responsibility.
4. **Resource Constraints:** Institutions may lack the necessary resources, expertise, or infrastructure to effectively monitor and regulate AI technologies.
5. **Bureaucratic Inertia:** Slow-moving institutions may struggle to keep pace with rapid technological advancements, hindering timely and effective regulation.

Overcoming these barriers requires coordination, investment in expertise, and a commitment to ethical standards across sectors and borders.

²⁷ Frahm et al. (2021); see also Irwin (2006).

The AoR wondered if he should read the bot's last sentence as a pragmatic policy blueprint, a cheerleading slogan, or an AI-induced hallucination. *Perhaps it was all three!*

22. *Overcoming these barriers requires coordination, investment in expertise, and a commitment to ethical standards across sectors and borders. Put this way, it sounds pretty easy. Is it?*

Implementing the kind of coordination and commitment needed to overcome barriers to regulating AI and chatbots is far from easy. While it sounds straightforward in theory, the reality involves navigating complex challenges such as aligning the interests of diverse stakeholders, overcoming political and economic resistance, managing international cooperation, and ensuring that regulations keep pace with rapidly evolving technologies. Each of these requires significant effort, resources, and ongoing vigilance, making the process difficult and fraught with potential setbacks. It's a multifaceted challenge that demands persistent and coordinated global action.

23. *Are there many precedents for successful achievement of this kind of "persistent and coordinated global action"?*

There are some precedents for achieving "persistent and coordinated global action," but they are relatively rare and often face significant challenges. Successful examples include international agreements on climate change, like the Paris Agreement, and global public health initiatives, such as the eradication of smallpox. However, these efforts typically involve long negotiation periods, varying levels of commitment from different countries, and ongoing challenges in implementation and enforcement. The complexities of aligning diverse interests and maintaining long-term cooperation make such global actions difficult to achieve and sustain.

24. *Well, the Paris Agreement looks like a mixed success. But let's set that aside. Part of the problem of reaching agreement is that technocratic policymaking cannot do the job. Generative AI raises problems that are as much political as technical. To take just one example, creating "diverse, representative training data" would require building agreement about what kinds of diversity matter. Do we only care about demographic diversity, or also about ideological diversity? Can a "representative" training set consist mainly of contemporary sources, or is that a form of historical imbalance? What baseline level of representation should be treated as "normal" and used to define under- and overrepresentation? Should any "fringe" viewpoints be excluded entirely? And who decides such questions? These issues are less about statistical representation than political representation. And people will continue to disagree about these and many other questions. The prospects for this changing are small, no? We're running out of time, so answer with a tweet.*

You're right; regulating generative AI involves deeply political questions about representation that go beyond technical fixes. Agreement on these issues is

unlikely, and the prospects for change are slim. #Alethics
#PoliticalRepresentation #AIregulation

25. *Finally, do you think that AI companies who have the capacity to implement their preferred solutions and do so without engaging in inclusive, transparent decision-making are violating principles of democracy?*

Yes, when AI companies implement their preferred solutions without engaging in inclusive, transparent decision-making, they are violating principles of democracy. Such actions sideline public input and accountability, which are essential for ensuring that technology serves the broader societal good rather than just the interests of a few.²⁸

Discussion

On reflecting on the results of this experiment, the AoR decided to begin the concluding discussion with a query:

26. *Write a 150-word commentary that reflects on the irony that a chatbot can now write a commentary on the institutional problems caused by the fact that chatbots can write commentaries.*

The irony of a chatbot discussing the institutional problems caused by chatbot-generated content is profound. As artificial intelligence develops the ability to critique its impact on human discourse, it exemplifies both a technological triumph and a paradox. Chatbots, by generating content on such issues, underscore the recursive nature of AI's role in modern communication. They can articulate problems regarding authenticity and intellectual integrity, yet they lack the self-awareness and depth to fully comprehend or resolve these issues. This scenario highlights the dual role of AI as both an analyzer and a contributor to the very challenges it identifies, reflecting the complex relationship between AI capabilities and their societal implications.

As is conventional, the AoR wrapped up the paper by briefly considering directions for future research.

27. *What sociopolitical problems associated with the recursive incorporation of large language models into societal discourse do you believe cannot now be anticipated? 50 words.*

Unforeseen sociopolitical problems from recursive incorporation of large language models may include the amplification of echo chambers, creation of

²⁸ The AoR noted that beyond sidelining public input, the transformation of societies into testbeds for LLM technology also imposes on a wide range of institutions the burden of attempting to formulate appropriate responses.

new forms of digital divides, unpredictable shifts in public trust and reliance on AI for information, and novel methods of censorship or control facilitated through AI-mediated communication.

28. *How can social research help address these "unforeseen" sociotechnical problems? 25 words.*²⁹

Social research can uncover patterns and consequences by analyzing data trends, monitoring societal responses, and fostering interdisciplinary collaborations to develop proactive measures and policies.

Finally, the AoR hopes this paper will inspire other researchers to reflect on the worldwide societal experiment with generative AI, perhaps by actively seeking help from LLM-based bots. Or perhaps by absorbing their assistance unknowingly, a possibility that grows increasingly likely as the use of AI increasingly infiltrates the modes of communication that shape the ways of knowing that constitute our world(s).

References

The chatbot provided no references for the sources it consulted.

Bibliography (by the Author-of-Record)

Ashmore, Malcolm. (1989). *The reflexive thesis: Wrighting sociology of scientific knowledge*. Chicago, IL: University of Chicago Press.

Barocas, Solon and Levy, Karen. (2020). Privacy dependencies. *Washington Law Review* 95(2).

Beck, Ulrich. (2006). Living in the world risk society. *Economy and Society* 35(3): 329-345.

Biagioli, Mario and Galison, Peter (Eds.). (2003). *Scientific Authorship: Credit and Intellectual Property in Science*. London: Routledge.

Beer, David. (2022). The problem of researching a recursive society: Algorithms, data coils, and the looping of the social. *Big Data & Society*, 9(2).

Borges, Jorge Luis. (1962). *Ficciones*. Edited and with an introduction by Anthony Kerrigan. New York, NY: Grove Press.

Boyle, James. (1996). *Shamans, Software, and Spleens: Law and the Construction of the Information Society*. Cambridge, MA: Harvard University Press.

²⁹ See Jasanoff's (2016, pp. 21-26) critique of "unintended consequences."

Cornell University Task Force on Generative AI in Academic Research. (2023). *Generative AI in Academic Research: Perspectives and Cultural Norms*. 15 December. <https://research-and-innovation.cornell.edu/generative-ai-in-academic-research/> (downloaded 12 October 2024).

Daston, Lorraine. (1992). Objectivity and the escape from perspective. *Social Studies of Science* 22(4): 597-618.

Daston, Lorraine and Galison, Peter. (2010). *Objectivity*. New York: Zone Books.

Doezema, Tess and Frahm, Nina. (2023). How the state built this AI moment. *The New Atlantis* 73(Summer): 46-53.

Edwards, Benj. (2023). ChatGPT sets record for fastest growing user base in history, report says. *Ars Technica* (1 February).

Engels, Franziska, Wentland, Alexander, and Pfothenhauer, Sebastian M. (2019). Testing future societies? Developing a framework for test beds and living labs as instruments of innovation governance. *Research Policy* 48: 103826.

Fortun, Mike. (2001). Mediated speculations in the genomics futures markets. *New Genetics & Society* (20)2: 139-156.

Foucault, Michel. (1975). What is an author? In Paul Rabinow (Ed.), *The Foucault Reader*, 101-120. New York: Pantheon.

Frahm, Nina, Doezeema, Tess, and Pfothenhauer, Sebastian. (2022). Fixing technology with society: The coproduction of democratic deficits and responsible innovation at the OECD and the European Commission. *Science, Technology, & Human Values* 47(1): 174-216.

Gieryn, Thomas F. (1999). *Cultural Boundaries of Science: Credibility on the Line*. Chicago, University of Chicago Press.

Gillespie, Tarleton. (2014). The relevance of algorithms. In Tarleton Gillespie, Pablo Boczkowski, & Kirsten Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167-195). Cambridge, MA: MIT Press.

Hacking, Ian. (2002). *Historical Ontology*. Cambridge, MA: Harvard University Press.

Hilgartner, Stephen. (2000). *Science on Stage: Expert Advice as Public Drama*. Stanford, CA: Stanford University Press.

Hilgartner, Stephen. (2008). Intellectual property and the politics of emerging technology: Inventors, citizens, and powers to shape the future. *Chicago-Kent Law Review* 84(1):197-224.

- Hilgartner, Stephen. (2017). *Reordering life: Knowledge and control in the genomics revolution*. Cambridge, MA: MIT Press.
- Hurlbut, J. Benjamin, Sheila Jasanoff, and Saha, Krishana. (2020). Constitutionalism at the nexus of life and law. *Science, Technology, & Human Values* 45(6) 979-1000.
- Irwin, Alan. (2006). The politics of talk: Coming to terms with the 'new' scientific governance. *Social Studies of Science* 36(2): 299-320.
- Jasanoff, Sheila. (2003). Technologies of humility: Citizen participation in governing science. *Minerva* 41(3).
- Jasanoff, Sheila. (2005). *Designs on nature: Science and democracy in Europe and the United States*. Princeton, NJ: Princeton University Press.
- Jasanoff, Sheila. (2016). *The ethics of innovation: Technology and the human future*. New York, NY: Norton.
- Jasanoff, Sheila. (2019). *Can science make sense of life?* New York: Wiley.
- Joly, Pierre-Benoit. (2015). Governing emerging technologies? The need to think outside the (black) box. In Stephen Hilgartner, Clark A. Miller, and Rob Hagendijk (Eds.), *Science and democracy: Making knowledge and making power in the biosciences and beyond*. New York: Routledge.
- Knorr Cetina, Karin. (1999). *Epistemic cultures: How the sciences make knowledge*. Cambridge, MA: Harvard University Press.
- Laurent, Brice and Tironi, Martin. (2015). A field test and its displacements. Accounting for an experimental mode of industrial innovation. *CoDesign* (11)3-4: 208-221. DOI: 10.1080/15710882.2015.1081241
- Marres, Noortje. (2020). What if nothing happens? Street trials of intelligent cars as experiments in participation. In Sabine Maasen, Sascha Dickel, Christoph Schneider (Eds.) *TechnoScienceSociety: Technological Reconfigurations of Science and Society*. Sociology of the Sciences Yearbook. Cham, Switzerland: Springer.
- Pfotenhauer, Sebastian, Laurent, Brice, Papageorgiou, Kyriaki, and Stilgoe, Jack. (2022). The politics of scaling. *Social Studies of Science* 52(1): 3-34.
- Parthasarathy, Shobita. (2017). *Patent politics: Life forms, markets, and the public interest in the United States and Europe*. Chicago: University of Chicago Press.
- Porter, Theodore. (1995). *Trust in numbers: the pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.

Rheinberger, Hans-Jörg. (1997). *Toward a history of epistemic things*. Stanford, CA: Stanford University Press.

Solaiman, Irene. (2023) "The gradient of generative AI release: Methods and considerations," February 5, 2023, <http://arxiv.org/pdf/2302.04844>

Stilgoe, Jack. (2018). Machine learning, social learning and the governance of self-driving cars. *Social Studies of Science* 48(1): 25–56.

Tennant, Chris and Stilgoe, Jack. (2021). The attachments of 'autonomous' vehicles. *Social Studies of Science* 51(6): 846–870.

Turkle, Sherry. (2005). *The second self: Computers and the human spirit (Twentieth anniversary edition)*. Cambridge, MA: MIT Press.

Vaidhyanathan, Siva. (2018). *Anti-social media*. New York: Oxford University Press.

Vertesi, Janet, & Ribes, David (Eds.). (2019). *digitalSTS: A field guide for science & technology studies*. Princeton, NJ: Princeton University Press.

Ziewitz, Malte. (2016). Governing algorithms: Myth, mess, and methods. *Science, Technology, & Human Values*, 41(1), 3-16.

Zuboff, Shoshana. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New York, NY: Public Affairs.